REPORT DOCUMENTATION PAGE

Form Approved OMB NO. 0704-0188

1. REPORT DATE (DD-MM-YYYY)	2. REPORT TYPE	3. DATES COVERED (From - To)
17-07-2010	Final Report	1-Sep-2009 - 31-May-2010

4. TITLE AND SUBTITLE
Data Representation: Learning Kernels from Noisy Data and Uncertain Information

5a. CONTRACT NUMBER
W911NF-09-1-0421

5b. GRANT NUMBER

5c. PROGRAM ELEMENT NUMBER
611102

6. AUTHORS
Rong Jin; Anil K. Jain

5d. PROJECT NUMBER

5e. TASK NUMBER

5f. WORK UNIT NUMBER

7. PERFORMING ORGANIZATION NAMES AND ADDRESSES
Michigan State University
Contract & Grant Admin.
Michigan State University
East Lansing, MI 48824 -

8. PERFORMING ORGANIZATION REPORT NUMBER

9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES)
U.S. Army Research Office
P.O. Box 12211
Research Triangle Park, NC 27709-2211

10. SPONSOR/MONITOR'S ACRONYM(S)
ARO

11. SPONSOR/MONITOR'S REPORT NUMBER(S)
56976-NS-II.1

12. DISTRIBUTION AVAILIBILITY STATEMENT
Approved for Public Release; Distribution Unlimited

13. SUPPLEMENTARY NOTES
The views, opinions and/or findings contained in this report are those of the author(s) and should not contrued as an official Department of the Army position, policy or decision, unless so designated by other documentation.

14. ABSTRACT
Identifying appropriate data representation is critical to many decision making problems. In this project, we focus on learning kernel-based data representation from noisy data and uncertain information. Unlike conventional approaches that represent objects by vectors, kernel representation defines a pairwise similarity

15. SUBJECT TERMS
Kernel Learning, learning from noisy information

16. SECURITY CLASSIFICATION OF:			17. LIMITATION OF ABSTRACT	15. NUMBER OF PAGES	19a. NAME OF RESPONSIBLE PERSON
a. REPORT	b. ABSTRACT	c. THIS PAGE			Rong Jin
UU	UU	UU	UU		19b. TELEPHONE NUMBER 517-353-7284

Report Title

Data Representation: Learning Kernels from Noisy Data and Uncertain Information

ABSTRACT

Identifying appropriate data representation is critical to many decision making problems. In this project, we focus on learning kernel-based data representation from noisy data and uncertain information. Unlike conventional approaches that represent objects by vectors, kernel representation defines a pairwise similarity between two objects, and is convenient for representing complex objects like graphs. Although many studies are devoted to learning kernel representation, none of them addresses the challenge of learning kernel representation from noisy data and uncertain information. The proposed research aims to address this challenging problem by developing (i) a kernel learning framework that are robust to data noise and information uncertainty, and (ii) efficient algorithms to solve the related optimization problems. The proposed algorithms will be evaluated in the object recognition domain. The impact of the proposed research to the US Army is significant. To counter against future threats to the safety and security of our society, we need to enhance our capabilities to detect, locate, and track such threats by extracting and representing data from noisy observation and uncertain information. The proposed research seeks to significantly advance, both theoretically and computationally, the representation and modeling of information from noisy and uncertain sources.

List of papers submitted or published that acknowledge ARO support during this reporting period. List the papers, including journal references, in the following categories:

(a) Papers published in peer-reviewed journals (N/A for none)

Number of Papers published in peer-reviewed journals: 0.00

(b) Papers published in non-peer-reviewed journals or in conference proceedings (N/A for none)

Number of Papers published in non peer-reviewed journals: 0.00

(c) Presentations

Number of Presentations: 0.00

Non Peer-Reviewed Conference Proceeding publications (other than abstracts):

Number of Non Peer-Reviewed Conference Proceeding publications (other than abstracts): 0

Peer-Reviewed Conference Proceeding publications (other than abstracts):

1. Learning from Noisy Side Information by Generalized Maximum Entropy Model, Tianbao Yang, Rong Jin, and Anil Jain, Proceedings of International Conference on Machine Learning (ICML), 2010

2. Unsupervised Transfer Classification: Application to Text Categorization, Tianbao Yang, Rong Jin, Anil K. Jain, Yang Zhou, Wei Tong, Proceedings of Knowledge Discovery and Data Mining (KDD), 2010

3. Online Kernel Learning, Rong Jin, Steven Hoi, and Tianbao Yang, Proceedings of Algorithmic Learning Theory (ALT), 2010

Number of Peer-Reviewed Conference Proceeding publications (other than abstracts): 3

(d) Manuscripts

Number of Manuscripts: 0.00

Patents Submitted

Patents Awarded

Graduate Students

NAME	PERCENT_SUPPORTED
Jinfeng Yi	1.00
FTE Equivalent:	**1.00**
Total Number:	**1**

Names of Post Doctorates

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Names of Faculty Supported

NAME	PERCENT_SUPPORTED	National Academy Member
Rong Jin	0.07	No
Anil K. Jain	0.03	No
FTE Equivalent:	**0.10**	
Total Number:	**2**	

Names of Under Graduate students supported

NAME	PERCENT_SUPPORTED
FTE Equivalent:	
Total Number:	

Student Metrics

This section only applies to graduating undergraduates supported by this agreement in this reporting period

The number of undergraduates funded by this agreement who graduated during this period:	0.00
The number of undergraduates funded by this agreement who graduated during this period with a degree in science, mathematics, engineering, or technology fields:.....	0.00
The number of undergraduates funded by your agreement who graduated during this period and will continue to pursue a graduate or Ph.D. degree in science, mathematics, engineering, or technology fields:.....	0.00
Number of graduating undergraduates who achieved a 3.5 GPA to 4.0 (4.0 max scale):	0.00
Number of graduating undergraduates funded by a DoD funded Center of Excellence grant for Education, Research and Engineering:	0.00
The number of undergraduates funded by your agreement who graduated during this period and intend to work for the Department of Defense	0.00
The number of undergraduates funded by your agreement who graduated during this period and will receive scholarships or fellowships for further studies in science, mathematics, engineering or technology fields:	0.00

Names of Personnel receiving masters degrees

NAME

Total Number:

Names of personnel receiving PHDs

NAME

Total Number:

Names of other research staff

NAME PERCENT_SUPPORTED

FTE Equivalent:

Total Number:

Sub Contractors (DD882)

Inventions (DD882)

Data Representation: Learning Kernels from Noisy Data and Uncertain Information
Proposal Number: 56976-NS
Rong Jin and Anil Jain, Michigan State University

Statement of Problem: Identifying appropriate data representation is critical to many problems in pattern recognition, data mining, and machine learning. Compared to the vector-based representation, kernel-based data representation is more flexible and is particularly suitable for complex objects like trees and graphs that are difficult to be captured by vector-based representation. In this project, we focus on the problem of automatically learning kernel-based data representation from noisy data and uncertain information. This is contrast to most current studies on kernel learning that assume an ideal observation or sensing of objects without any noise. The objective of this project is to develop efficient computational frameworks for learning a robust combination of multiple kernel data representations from noisy data observations and uncertain supervised information. The proposed research aims to develop the following approaches to address the key challenges in kernel learning with noisy data and uncertain information

1. Develop an efficient computational framework for multiple kernel learning that is resilient to the noise in data observation
2. Develop an efficient computational framework for multiple kernel learning that is robust to the uncertainty in class assignment

Significance: This project addresses one fundamental problem in pattern recognition and machine learning, i.e., how to derive accurate data representation from noisy observation and uncertain side information. The result of this result will lead to significant progress in kernel learning, a critical component to many pattern recognition and machine learning algorithms and theories. Given the growing threats of global terrorism and illegal activities such as transportation of hazardous materials and human trafficking, the result of this research will significantly advance, both theoretically and computationally, the representation and modeling of information from noisy and uncertain sources, which in return improves our capabilities to detect, locate, and track various threats. The results of this research will benefit the Army by expanding the wealth of information that can be utilized in a network-centric environment to support effective and reliable decision making during combat missions and in the global war on terrorism. The theoretical and computational advances proposed in this project are also important key steps toward enabling the Army to gain information superiority, which is crucial to ensure the success of its future missions.

Summary of the Most Important Results:

1. Online kernel learning. Although a large number of studies are devoted to kernel learning, most of them suffer from the high computational cost, making them inefficient for handling a number of training examples. We address this challenge by developing an online learning theory for kernel learning. Compared to the existing approaches, online kernel learning is computationally more efficient as it only needs to scan the entire set of training examples once. Online kernel learning is generally more challenging than typical online learning because it requires learning both the kernel classifiers and their combination weights simultaneously. We have developed both deterministic approaches and stochastic approaches for online kernel learning. We have derived mistake bounds for both algorithms. This work has been published in the proceeding of Algorithmic Learning Theory (ALT) 2010 [1].

2. Kernel learning from noisy side information. Most studies on kernel learning assume that the side information, such as pairwise constraints, is perfect without any error. In this project, we examine the problem of kernel learning from noisy side information in the form of pairwise constraints. We emphasize that this is an important problem because pairwise constraints are often extracted from data sources such as paper citations, and therefore are usually noisy and inaccurate. To address this challenging problem, we introduce the Generalized Maximum Entropy (GME) model and propose a framework for learning a combination of kernels from noisy side information based on the GME model. The theoretic analysis shows that under appropriate assumptions, the classification model trained from the noisy side information can be very close to the one trained from the perfect side information. Extensive empirical studies verify the effectiveness of the proposed framework. This work has been published in International Conference on Machine Learning (ICML) 2010 [2].

3. Unsupervised kernel classification. We study the problem of building the kernel classifier for a target class in the absence of any labeled training example for that class. To address this difficult learning problem, we extend the idea of transfer learning by assuming that the following side information is available: (i) a collection of labeled examples belonging to other classes in the problem domain, called the auxiliary classes; (ii) the class information including the prior of the target class and the correlation between the target class and the auxiliary classes. Our goal is to construct a kernel classifier for the target class by leveraging the above data and information. Our framework is based on the generalized maximum entropy model that is effective in transferring the label information of the auxiliary classes to the target class. A theoretical analysis shows that under certain assumption, the classification model obtained by the proposed approach converges to the optimal model when it is learned from the labeled examples for the target class. Empirical study on text categorization over four different data sets verifies the effectiveness of the proposed approach. This work has been published in ACM Conference on Knowledge Discovery and Data Mining (KDD), 2010

Bibliography:

[1]. Online Kernel Learning, Rong Jin, Steven Hoi, and Tianbao Yang, Proceedings of Algorithmic Learning Theory (ALT), 2010

[2] Learning from Noisy Side Information by Generalized Maximum Entropy Model, Tianbao Yang, Rong Jin, and Anil Jain, Proceedings of International Conference on Machine Learning (ICML), 2010

[3] Unsupervised Transfer Classification: Application to Text Categorization, Tianbao Yang, Rong Jin, Anil K. Jain, Yang Zhou, Wei Tong, Proceedings of Knowledge Discovery and Data Mining (KDD), 2010